[P5-07-01] Successful whole transcriptome analysis of 25-year-old breast tumor samples from the phase III trial SWOG-8814 by next generation sequencing (NGS): Standardized analytical methods for exploratory and validation studies

Cherbavaz DB, Hayes DF, Qu K, Crager MR, Barlow WR, Goddard AD, Beasley EM, Jeong J, Collin F, Liu M-L, Rae JM, Ravdin PM, Tripathy D, Gralow JR, Livingston RB, Osborne CK, Ingle JN, Pritchard KI, Davidson NE, Carey LA, Sing AP, Baehner FL, Hortobagyi GN, Shak S, Albain KS. Genomic Health, Inc., Redwood City, CA; Univerity of Michigan, Ann Arbor, MI; Cancer Research and Biostatistics, Seattle, WA; Cancer Therapy & Research Center (CTRC) of The University of Texas Health Science Center, San Antonio, TX; The University of Texas MD Anderson Cancer Center, Houston, TX; University of Washington, Seattle Cancer Care Alliance, Seattle, WA; Univeristy of Arizona Cancer Center, Tuscon, AR; Baylor College of Medicine, Houston, TX; Mayo Clinic, Rochester, MN; Sunnybrook Odette Cancer Centre and the University of Toronto, Toronto, ON, Canada; University of Pittsburgh Cancer Center, Pittsburgh, PA; University of North Carolina at Chapel Hill, Chapel Hill, NC; Genomic Health, Inc. and University of California, San Francisco, Redwood City and San Francisco, CA; Loyola University Chicago Stritch School of Medicine, Maywood, IL

BACKGROUND: We previously reported that low 21 gene Recurrence Scores (RS) identify patients with ER-positive, lymph node-positive breast cancer who may not benefit from anthracycline-based adjuvant chemotherapy added to tamoxifen (SWOG-8814A NCI correlative science study; Albain et al. Lancet Oncol 2010). New exploratory and comprehensive quantitative analyses now permit whole transcriptome NGS on residual RNA extracted from FFPE blocks 12-18 years post-fixation. Herein, we report methodology details and feasibility results (see companion abstract, Albain et al., for clinical outcomes correlations).
METHODS and RESULTS: Sequencing was carried out in Illumina HiSeq 2000 instruments, yielding 4.2 trillion data points. Messenger RNA expression was quantified using 3rd quartile normalization. Both Library (RNA) and Sequencing Standards showed high quality coverage as measured by median uniquely mapped reads over a 13 month window (168M and 182M, respectively, including duplicate reads). The median absolute deviation (MAD) of the relative log expression (RLE) of mapped reads for the Library and Sequencing Standards was 0.22 and 0.05, respectively. The Library Standard variation was greater than the Sequencing Standard, as library preparation was manual. Of 360 patient samples with sufficient RNA (≥ 100 ng total RNA), 354 (98.3%) were successfully sequenced and included in the final analysis data set. Average library yield was 39 ng/μL. Only 5 libraries failed yield requirements and one library failed expression quality metrics. The median insert length was 120 bp with the first and third quartiles 93 and 152 bp, respectively. After removal of duplicate reads, 82% of reads were uniquely mapped, and the median library size was 8.95M (number of unique mapped reads). Sequences with counts <10 for all 354 patients were excluded. The medians of the 1st, 2nd and 3rd quartiles for exons mapped to the RefSeq database were 20, 40 and 78 counts, and for introns, 72, 134 and 244 counts, respectively. The majority of exonic (86.7%) and intronic (95%) sequences were mapped. There were 20,101 RefSeq mapped exons with counts ≥10. Of these exons, 988 passed additional filtering criteria and were subjected to hierarchical clustering, with Gene Ontology and pathway analysis performed on selected gene expression patterns (for results, see companion abstract of Albain et al.).
CONCLUSIONS: High quality whole transcriptome NGS is feasible from decades-old clinical trial FFPE specimens that have not been stored in any special fashion. Controlled laboratory, bioinformatics and biostatistics methods, with inclusion of appropriate process controls, ensure robustness and reliability of the NGS process. This in turn results in the discovery and validation of biologically and clinically relevant variations from prior landmark clinical trials.
SUPPORT: NCI CA 180888, CA180819, CA180821, CA180820, CA180863; in part, Genomic Health, Inc.

Friday, December 11, 2015 5:00 PM

Poster Session 5: Prognostic and Predictive Factors: Biomarkers -- Methods (5:00 PM-7:00 PM)

Terms of Service.

Close Window